黄侧异腹胡蜂线粒体基因组全序列测定和分析

彭　艳, 陈　斌, 李廷景*

（重庆师范大学昆虫与分子生物学研究所, 动物生物学重庆市重点实验室, 媒介昆虫重庆市重点实验室, 重庆 401331）

摘要:【目的】测序和分析黄侧异腹胡蜂 *Parapolybia crocea* 线粒体基因组,并在线粒体基因组水平探讨异腹胡蜂属 *Parapolybia* 在胡蜂科中的系统发育地位。【方法】用 Illumina 二代测序技术测定黄侧异腹胡蜂线粒体基因组全序列,分析其结构特点和碱基组成;使用最大似然法(maximum likelihood, ML)构建胡蜂科 7 个种线粒体基因组的系统发育树,分析其在胡蜂科中的系统发育关系。【结果】黄侧异腹胡蜂线粒体基因组全长 16 619 bp(GenBank 登录号: KY679828),包含 13 个蛋白质编码基因,22 个 tRNA 基因,2 个 rRNA 基因(*rrnS* 和 *rrnL*)和 1 个控制区,基因排列顺序与推测的昆虫祖先序列不完全一致;全部蛋白质编码基因的起始密码子均为 ATN,终止密码子除 *CYTB* 和 *ND*1 为 TAG 外,其余均为 TAA;除 tRNA$^{Ser(AGN)}$ 的 DHU 臂缺失外,其他 tRNA 均能折叠成典型的三叶草结构;控制区中存在一个 18 bp 的 T-stretch 结构和 2 段串联重复序列。胡蜂科 7 个种基于线粒体基因组的系统发育关系表现为蜾蠃亚科 + (胡蜂亚科 + 马蜂亚科),异腹胡蜂属与马蜂属 *Polistes* 同属于马蜂亚科。【结论】黄侧异腹胡蜂线粒体基因组存在基因重排现象。基于线粒体基因组的胡蜂科系统发育关系与传统的形态分类学结果一致:异腹胡蜂属隶属于马蜂亚科,马蜂亚科与胡蜂亚科的亲缘关系较其与蜾蠃亚科更近。

关键词: 膜翅目; 胡蜂科; 异腹胡蜂属; 黄侧异腹胡蜂; 线粒体基因组; 系统发育关系

中图分类号: Q966　　**文献标识码:** A　　**文章编号:** 0454-6296(2017)04-0464-11

Sequencing and analysis of the complete mitochondrial genome of *Parapolybia crocea* (Hymenoptera: Vespidae)

PENG Yan, CHEN Bin, LI Ting-Jing* (Chongqing Key Laboratory of Vector Insects, Chongqing Key Laboratory of Animal Biology, Institute of Insect and Molecular Biology, Chongqing Normal University, Chongqing 401331, China)

Abstract:【Aim】To sequence and analyze the complete mitochondrial genome of *Parapolybia crocea*, and to explore the phylogenetic position of *Parapolybia* in Vespidae based on mitochondrial genome sequences.【Methods】The complete mitochondrial genome of *P. crocea* was sequenced by using Illumina sequencing technique, and its general features and base composition were analyzed. The phylogenetic tree of mitochondrial genomes of seven vespid species was constructed by using maximum likelihood method, and their phylogenetic relationships within Vespidae were analyzed.【Results】The mitochondrial genome of *P. crocea* is a circular molecule of 16 619 bp (GenBank no.: KY679828), including 13 protein coding genes, 22 tRNA genes, two rRNA genes (*rrnS* and *rrnL*), and one control region (CR). The gene arrangement of *P. crocea* is not completely consistent with the putative ancestral sequence of insects. All protein-coding genes initiate with the start codon ATN. The stop codon of these genes is TAA, except for *CYTB* and *ND*1 which terminates with TAG. All the predicted tRNAs show the classic clover-leaf structure, except for the absence of DHU arm of tRNA$^{Ser(AGN)}$. The control region contains one 18 bp T-stretch structure and two tandem repeat sequences. The phylogenetic relationships

基金项目: 国家自然科学基金项目(31372247, 31000976, 31672363, 31372265); 重庆师范大学青年拔尖人才支持项目(14CSDG07); "两江学者"计划项目

作者简介: 彭艳, 女, 1993 年 11 月生, 重庆潼南人, 硕士研究生, 研究方向为昆虫学, E-mail: py1993@yeah.net

* 通讯作者 Corresponding author, E-mail: ltjing1979@hotmail.com

收稿日期 Received: 2017-01-27; 接受日期 Accepted: 2017-03-19

of seven species within Vespidae based on the mitochondrial genomes present as Eumeninae + (Vespinae + Polistinae), and *Parapolybia* and *Polistes* are grouped into the subfamily Polistinae. 【Conclusion】 Gene rearrangement exists in the mitochondrial genome of *P. crocea*. The phylogenetic relationships of Vespidae constructed based on the mitochondrial genomes are consistent with those in traditional morphological taxonomy: *Parapolybia* is grouped into Polistinae, and the relationship between Polistinae and Vespinae is closer than that between Polistinae and Eumeninae.

Key words: Hymenoptera; Vespidae; *Parapolybia*; *Parapolybia crocea*; mitochondrial genome; phylogentic relationship

动物线粒体基因组是一个大小为 15 ~ 18 kb 的双链闭合环状分子(Wolstenholme, 1992; Boore, 1999; Cameron, 2014),具有结构简单、基因重组极少、母系遗传和进化速率较快等特点(Song *et al*., 2016b),是系统发育和种群遗传研究的重要分子标记。膜翅目(Hymenoptera)线粒体基因组除了上述特征外,还具有 A + T 含量通常较其他目高(Dowton and Austin, 1997; Cameron *et al*., 2008; Wei, 2010),且基因重排现象较明显的特点。如细腰亚目(Apocrita)的重排率高于广腰亚目(Symphyta)(Jermiin and Crozier, 1994),重排事件主要发生在 tRNA 基因。与昆虫纲其他类群线粒体基因组相似,膜翅目线粒体基因组编码 37 个基因,包括 13 个蛋白质编码基因,22 个 tRNA 基因和 2 个 rRNA 基因,还有 1 个富含 AT 的控制区(Boore, 1999)。

异腹胡蜂属 *Parapolybia* 昆虫是社会性昆虫,常在树干、墙壁、房缘等地方筑巢,主要以鳞翅目(Lepidoptera)昆虫的幼虫为食,是一类重要的天敌昆虫。该属创建于 1854 年(Saussure, 1853 - 1858),Richards(1962)将其归为马蜂亚科 Polybiini 族(Carpenter, 1993),我国学者李铁生(1982, 1985)记述为异腹胡蜂科,直到 Carpenter(1993)通过形态特征对胡蜂科分类系统进行探讨,将异腹胡蜂属归为马蜂亚科 Ropalidiini 族,并沿用至今。目前,该属尚无线粒体基因组序列报道,整个胡蜂科也仅报道了 6 种线粒体基因组:马蜂亚科 2 种、胡蜂亚科 3 种及蜾蠃亚科 1 种(表 1)。

表 1　已报道线粒体基因组的胡蜂科种类

Table 1　Species with their mitochondrial genomes known in Vespidae

亚科 Subfamily	种 Species	GenBank 登录号 GenBank accession no.	参考文献 References
马蜂亚科 Polistinae	黄侧异腹胡蜂 *Parapolybia crocea*	KY679828	本文 This paper
	矮马蜂 *Polistes humilis*	EU024653	Cameron *et al*., 2008
	约马蜂 *Polistes jokahamae*	KR052468	Song *et al*., 2016a
胡蜂亚科 Vespinae	双色胡蜂 *Vespa bicolor*	KJ735511	Wei *et al*., 2014
	金环胡蜂 *Vespa mandarinia*	KR059904	Chen *et al*., 2015
	德国黄胡蜂 *Vespula germanica*	KR703583	Zhou *et al*., 2016
蜾蠃亚科 Eumeninae	扁平黄蜾蠃 *Abispa ephippium*	NC011520	Cameron *et al*., 2008

黄侧异腹胡蜂 *Parapolybia crocea* 主要分布于中国(贵州、广东、广西、陕西、江西、湖北、湖南、安徽、福建和云南等)、日本、泰国、老挝和越南。本研究测序和分析黄侧异腹胡蜂线粒体基因组,在报道该种线粒体基因组特点的同时,从线粒体基因组水平探讨异腹胡蜂属在胡蜂科中的分类地位,并结合已报道的胡蜂科 6 个种线粒体基因组全序列分析各亚科之间的系统发育关系,为后续从线粒体基因组水平研究胡蜂科各亚科、各属及种类间的系统发育关系提供研究基础。

1　材料与方法

1.1　材料

黄侧异腹胡蜂 *Parapolybia crocea* 标本于 2015 年 8 月 11 日采集自重庆市云阳县农坝镇(31.22°N, 108.44°E,海拔 673 m),采集地的生态环境主要以农田为主,属高寒地区,雨量适中,常年平均气温

18℃左右。采集后的标本保存于无水酒精,储存在 -20℃冰箱中。

1.2 黄侧异腹胡蜂总 DNA 提取

总 DNA 提取自黄侧异腹胡蜂标本的胸部肌肉,采用 DNeasy DNA Extraction Kit 试剂盒(QIAGEN),参照说明书进行总 DNA 的抽提,提取完毕后,用双链 DNA 超敏试剂盒(Invitrogen)中的 Qubit 荧光计检测总 DNA 质量。

1.3 序列测定

本研究采用集合多物种 DNA 混合建库的方式,通过高通量测序技术完成测序工作(贝瑞和康)。取每个样品 5.0 μg 的基因组 DNA 构建 Illumina TruSeq 文库,平均插入片段长度为 480 bp,用 Illumina Hiseq 2500 对文库进行双端测序,平均测序深度为 50X,保证总数据量为 4 G(250 bp Reads)。移除连接物和不配对、短小、质量差的 Reads 后,将高质量的 Reads 用 IDBA-UD 进行从头组装(Peng *et al*., 2012)。

1.4 序列识别

参考 Simon 等(2006)设计的引物,通过标准 PCR 反应扩增各种线粒体基因组蛋白质编码基因 *COX*1 和 rRNA 基因 *rrnS* 的基因片段,作为靶标序列对线粒体基因组全序列进行 BLASTN 查找,配对率达到 100% 才确认为黄侧异腹胡蜂基因组。对线粒体基因组序列相同或接近相同的重叠端区域采用在线软件 Geneious(http://www.geneious.com/)进行人工查验,并在可能的情况下将其环化。此外,采用 PCR 扩增目的条带,切胶回收后克隆测序的方法对黄侧异腹胡蜂的控制区序列进行验证。

1.5 序列注释及特征分析

以已知的马蜂亚科线粒体基因组作为参考,用在线软件 tRNAscan-SE 1.0(http://lowelab.ucsc.edu/tRNAscan-SE)预测能够生成的 tRNA 基因,通过人工比对来注释蛋白质编码基因以及该软件无法识别的 tRNA 基因、rRNA 基因和控制区。使用 MEGA 6(Tamura *et al*., 2013)计算基因的核苷酸组成,包括 AT 含量、AT 偏斜、GC 偏斜,并分析蛋白质编码基因的氨基酸含量和同义密码子相对使用度(relative synonymous codon usage, RSCU)。tRNA 基因的二级结构同样用 tRNAscan-SE 1.0 进行预测,无法预测的则通过人工校正完成。用软件 Tandem Repeats Finder 4.09 对控制区的串联重复序列进行查找,绘制控制区的结构图。

1.6 系统发育分析

以与胡蜂科亲缘关系较近的蚁科黑火蚁 *Solenopsis richteri* 为外群(Brothers, 1999; Gotzek, 2010; Song *et al*., 2016a),选用 NCBI 中可获得的胡蜂科 6 个种和本研究中测得的黄侧异腹胡蜂线粒体基因组,基于 13 个蛋白质编码基因的核苷酸序列构建胡蜂科的系统发育树。将胡蜂科的 7 个种和 1 个外群线粒体基因组的 13 个蛋白质编码基因序列串联,用 Clustal X 进行比对,分别采用 PAUP 4.0 和 MEGA 6 软件预测其建树的最适模型,均选择最大似然法(maximum likelihood, ML)用 Bootstrap(1 000 次重复)检验各分支节点的置信值,构建两种 ML 树。

2 结果

2.1 黄侧异腹胡蜂线粒体基因组结构

黄侧异腹胡蜂的线粒体基因组全长 16 619 bp(GenBank 登录号:KY679828),与昆虫纲部分昆虫类似(Wolstenholme 1992; Cha *et al*., 2007; Zhao *et al*., 2015; Luo *et al*., 2016),基因组共包括 37 个基因(表 2):13 个蛋白质编码基因、22 个 tRNA 基因和 2 个 rRNA 基因(*rrnS* 和 *rrnL*),还有 1 个控制区。其中,14 个基因在 N 链上编码:4 个蛋白质编码基因、8 个 tRNA 基因和 2 个 rRNA 基因,其余 23 个基因在 J 链上编码。控制区长度为 1 316 bp,位于 *rrnS* 基因与 $tRNA^{Tyr}$ 基因之间。黄侧异腹胡蜂的线粒体基因排列顺序与昆虫假想的祖先序列不完全一致(Clary and Wolstenholme, 1985; Hua *et al*., 2016),存在基因重排现象(图 1),例如 $tRNA^{Tyr}$ 基因从 $tRNA^{Cys}$ 基因和蛋白质编码基因 *COX*1 之间重排到了控制区的下游和 $tRNA^{Ile}$ 基因的上游;蛋白质编码基因 *ND*1 与 $tRNA^{Leu(CUN)}$ 基因的位置发生了互换,形成基因簇 *trnS*2-*trnL*1-*ND*1。

黄侧异腹胡蜂线粒体基因组中,有 7 处基因重叠,共 23 bp,其中最长的一处重叠序列为 8 bp,位于 $tRNA^{Trp}$ 基因和 $tRNA^{Cys}$ 基因之间,其余长度为 1 ~ 7 bp;基因之间的间隔区域有 21 处,共 643 bp,长度为 1 ~ 135 bp 不等,其中 $tRNA^{Met}$ 基因和 *ND*2 之间的基因间隔最长;共有 9 处的基因直接相邻。

2.2 黄侧异腹胡蜂线粒体基因组核苷酸组成

黄侧异腹胡蜂线粒体基因组中,A 碱基占 43.4%,T碱基占39.6%,C碱基占11.1%,G碱基

表 2　黄侧异腹胡蜂线粒体基因组组成

Table 2　Composition of the mitochondrial genome of *Parapolybia crocea*

基因 Gene	方向 Direction	起始位置 Start position	终止位置 Stop position	大小(bp) Size	反密码子 Anticodon	起始密码子 Start codon	终止密码子 Stop codon	与下一基因的间隔(bp) Length to the next gene
$tRNA^{Ile}$	F	1	65	65	GAT			27
$tRNA^{Gln}$	F	80	146	67	TTG			14
$tRNA^{Met}$	F	148	215	68	CAT			1
*ND*2	F	351	1 274	924		ATA	TAA	135
$tRNA^{Trp}$	F	1 274	1 352	79	TCA			−1
$tRNA^{Cys}$	R	1 345	1 410	66	GCA			−8
*COX*1	F	1 414	2 952	1 539		ATG	TAA	3
$tRNA^{Leu(UUR)}$	F	2 971	3 039	69	TAA			18
*COX*2	F	3 040	3 729	690		ATT	TAA	0
$tRNA^{Lys}$	F	3 731	3 799	69	CTT			1
$tRNA^{Asp}$	F	3 837	3 903	67	GTC			37
*ATP*8	F	3 904	4 071	168		ATT	TAA	0
*ATP*6	F	4 079	4 741	663		ATG	TAA	7
*COX*3	F	4 780	5 565	786		ATG	TAA	38
$tRNA^{Gly}$	F	5 568	5 635	68	TCC			2
*ND*3	F	5 636	5 992	357		ATT	TAA	0
$tRNA^{Ala}$	F	5 997	6 062	66	TGC			4
$tRNA^{Arg}$	F	6 062	6 130	69	TCG			−1
$tRNA^{Asn}$	F	6 148	6 215	68	GTT			17
$tRNA^{Ser(AGN)}$	F	6 213	6 274	62	TCT			−3
$tRNA^{Glu}$	F	6 281	6 347	67	TTC			6
$tRNA^{Phe}$	R	6 350	6 422	73	GAA			2
*ND*5	R	6 463	8 133	1671		ATT	TAA	40
$tRNA^{His}$	R	8 178	8 241	64	GTG			44
*ND*4	R	8 274	9 590	1 317		ATG	TAA	32
ND4L	R	9 584	9 865	282		ATT	TAA	−7
$tRNA^{Thr}$	F	9 923	9 986	64	TGT			57
$tRNA^{Pro}$	R	9 986	10 053	68	TGG			−1
*ND*6	F	10 089	10 616	528		ATG	TAA	35
CYTB	F	10 617	11 759	1 143		ATG	TAG	0
$tRNA^{Ser(UCN)}$	F	11 758	11 825	68	TGA			−2
$tRNA^{Leu(CUN)}$	R	11 841	11 905	65	TAG			15
*ND*1	R	12 014	12 967	954		ATT	TAG	108
rrnL	R	12 968	14 391	1 424				0
$tRNA^{Val}$	R	14 392	14 460	69	TAC			0
rrnS	R	14 461	15 212	752				0
控制区 Control region		15 213	16 528	1 316				0
$tRNA^{Tyr}$	R	16 529	16 593	65	GTA			0

F：正向 Forward；R：反向 Reverse.

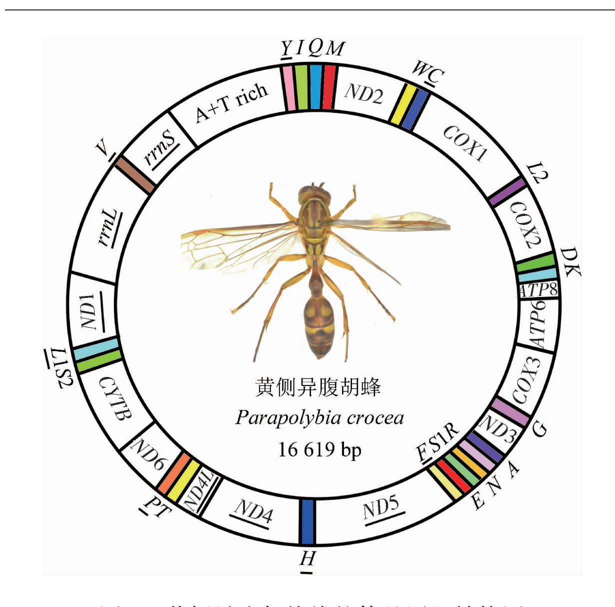


图 1 黄侧异腹胡蜂线粒体基因组结构图

Fig. 1 Structure of the mitochondrial genome of *Parapolybia crocea*

加下划线的基因表示在 N 链上编码，其余基因在 J 链上编码；填充的颜色框表示不同的 tRNA 基因，用对应的氨基酸单字母表示；未填充颜色的白色框分别表示蛋白质编码基因、rRNA 基因和控制区。The genes underlined are located on the N-strand and others on the J-strand. The color-filled blocks denote tRNA genes, which are indicated by one letter amino acid code of the corresponding amino acid. The white blocks denote protein-coding genes, rRNA genes and the control region, respectively.

占 5.9%，全基因组的 A + T 含量为 82.9%，G + C 含量为 17.1%，表现出明显的对 A 和 T 碱基的偏向性（表 3）。蛋白质编码基因全长 11 022 bp，A + T 含量为 80.6%，G + C 含量为 19.4%；代表氨基酸的密码子第 3 位碱基使用 A/T 碱基的频率为 90.8%，明显高于第 1 位的 77.4% 和第 2 位的 73.7%。密码子的第 3 位碱基大多为 A 碱基或 T 碱基的现象在膜翅目线粒体基因组中较为常见（Song *et al.*, 2016b）。22 个 tRNA 基因序列总全长为 1 486 bp，A + T含量为 86.1%。2 个 rRNA 基因全长为 2 176 bp，A + T 含量为 86.3%。控制区序列长度为 1 316 bp，A + T 含量为 88.3%。

黄侧异腹胡蜂线粒体基因组全序列的 AT 偏斜为正值（0.046），GC 偏斜为负值（－0.307），表明基因组序列中 A 碱基和 C 碱基的百分比含量高于 T 碱基和 G 碱基，与大部分昆虫的线粒体基因组 AT/GC 偏斜一致（Wei *et al.*, 2010）。蛋白质编码基因的 AT 偏斜为负（－0.120%）、GC 偏斜也为负（－0.015%），其中，位于 J 链上的 9 个蛋白质编码基因的 AT 偏斜和 GC 偏斜均为负，N 链编码的 4 个蛋白质基因 AT 偏斜为负，tRNA 基因 AT 偏斜为正，*rrnL* 和 *rrnS* 基因 AT 偏斜为负，GC 偏斜均为正值（表 3）。

表 3 黄侧异腹胡蜂线粒体基因组核苷酸组成

Table 3 Nucleotide composition of the complete mitochondrial genome of *Parapolybia crocea*

基因序列 Gene sequences	T (%)	C (%)	A (%)	G (%)	A + T (%)	G + C (%)	AT 偏斜 AT-skew	GC 偏斜 GC-skew
全基因组 Whole genome	39.6	11.1	43.4	5.9	82.9	17.1	0.046	－0.307
蛋白质编码基因 Protein-coding genes	45.2	9.8	35.5	9.5	80.6	19.4	－0.120	－0.015
密码子第 1 位 1st codon site	37.9	9.1	39.5	13.5	77.4	22.6	0.021	0.194
密码子第 2 位 2nd codon site	51.3	15.0	22.4	11.3	73.7	26.3	－0.393	－0.143
密码子第 3 位 3rd codon site	46.2	5.3	44.6	3.9	90.8	9.2	－0.019	－0.160
tRNA 基因 tRNA genes	42.1	6.2	44.0	7.7	86.1	13.9	0.022	0.107
rRNA 基因 rRNA genes	46.0	4.4	40.3	9.4	86.3	13.7	－0.066	0.365
控制区 Control region	46.0	6.5	42.2	5.2	88.3	11.7	－0.043	－0.117

2.3 黄侧异腹胡蜂线粒体基因组蛋白质编码基因

本研究中黄侧异腹胡蜂线粒体基因组蛋白质编码基因均以 ATN 作为起始密码子，其中 *ND2* 使用 ATA 作为起始密码子，*COX1*，*ATP6*，*COX3*，*ND4*，*ND6* 和 *CYTB* 使用 ATG 作为起始密码子，其余 6 个基因均使用 ATT 作为起始密码子。除了 *CYTB* 和 *ND1* 以 TAG 作为终止密码子，其余编码基因的终止密码子为 TAA，与昆虫线粒体基因组普遍使用的终止密码子为 TAA 或 TAG 相一致（Wolstenholme, 1992）。

除去终止密码子，黄侧异腹胡蜂线粒体蛋白质编码基因共包含 3 661 个密码子，其在昆虫线粒体蛋白质编码基因密码子数量在 3 585 ~ 3 746 范围内（Cha *et al.*, 2007）。相对同义密码子使用情况显示，其线粒体基因组对 A 和 T 碱基具有明显的偏向性（表 4），其中最常用的 4 个密码子是 UUU（Phe），UUA（Leu），AUU（Ile）和 AUA（Met），密码子使用次

数依次为 346，486，431 和 302，其第 3 位碱基均为 A 或 U，说明黄侧异腹胡蜂线粒体蛋白质编码基因密码子第 3 位偏好使用 A 或 U 碱基。蛋白质编码基因的氨基酸百分比含量最高为 Leu（15.7%），其次分别为 Ile（13.3%），Phe（10.3%）和 Met（8.8%），是使用频率最高的 4 个氨基酸（图 2），这一现象在其他膜翅目昆虫中也很常见（Wei *et al.*，2009）。

表 4 黄侧异腹胡蜂线粒体基因组密码子使用情况

Table 4 Codon usage in the mitochondrial genome of *Parapolybia crocea*

氨基酸 Amino acid	密码子 Codon	使用次数 n	同义密码子相对使用度 RSCU	氨基酸 Amino acid	密码子 Codon	使用次数 n	同义密码子相对使用度 RSCU
Phe（F）	UUU	346	1.84	Tyr（Y）	UAU	144	1.69
	UUC	31	0.16		UAC	26	0.31
Leu（L）	UUA	486	5.06	Stop codon*	UAA	11	1.69
	UUG	35	0.36		UAG	2	0.31
	CUU	24	0.25	His（H）	CAU	54	1.69
	CUC	3	0.03		CAC	10	0.31
	CUA	28	0.29	Gln（Q）	CAA	48	1.81
	CUG	0	0		CAG	5	0.19
Ile（I）	AUU	431	1.77	Asn（N）	AAU	235	1.88
	AUC	55	0.23		AAC	15	0.12
Met（M）	AUA	302	1.88	Lys（K）	AAA	134	1.9
	AUG	19	0.12		AAG	7	0.1
Val（V）	GUU	63	2	Asp（D）	GAU	47	1.74
	GUC	0	0		GAC	7	0.26
	GUA	56	1.78	Glu（E）	GAA	68	1.77
	GUG	7	0.22		GAG	9	0.23
Ser（S）	UCU	93	1.83	Cys（C）	UGU	27	1.8
	UCC	9	0.18		UGC	3	0.2
	UCA	100	1.97	Trp（W）	UGA	68	1.72
	UCG	1	0.02		UGG	11	0.28
Pro（P）	CCU	57	1.92	Arg（R）	CGU	14	1.3
	CCC	13	0.44		CGC	0	0
	CCA	48	1.61		CGA	26	2.42
	CCG	1	0.03		CGG	3	0.28
Thr（T）	ACU	58	1.48	Ser（AGN）	AGU	22	0.91
	ACC	14	0.36		AGC	1	0.04
	ACA	83	2.11		AGA	65	2.68
	ACG	2	0.05		AGG	9	0.37
Ala（A）	GCU	39	2.14	Gly（G）	GGU	45	1.09
	GCC	7	0.38		GGC	2	0.05
	GCA	26	1.42		GGA	88	2.13
	GCG	1	0.05		GGG	30	0.73

n：密码子使用次数 Frequency of a codon used；RSCU：同义密码子相对使用度 Relative synonymous codon usage；* 终止密码子 Stop codon.

2.4 黄侧异腹胡蜂线粒体基因组 tRNA 基因和 rRNA 基因

本研究的黄侧异腹胡蜂线粒体基因组具 20 个 tRNA 基因的反密码子，且与多数昆虫线粒体基因组对应 tRNA 的反密码子相同（Lee *et al.*，2008；王菊平等，2015；Hua *et al.*，2016），只有 $tRNA^{Ile}$ 的反密码子由 CCT 变为 GAT，$tRNA^{Ser(AGN)}$ 的反密码子由 GCT 变为 TCT，同时这一现象也存在于目前已报道的胡蜂科其他种类中（Cameron *et al.*，2008；Wei，2014；Chen，2015；Song *et al.*，2016a；Zhou *et al.*，

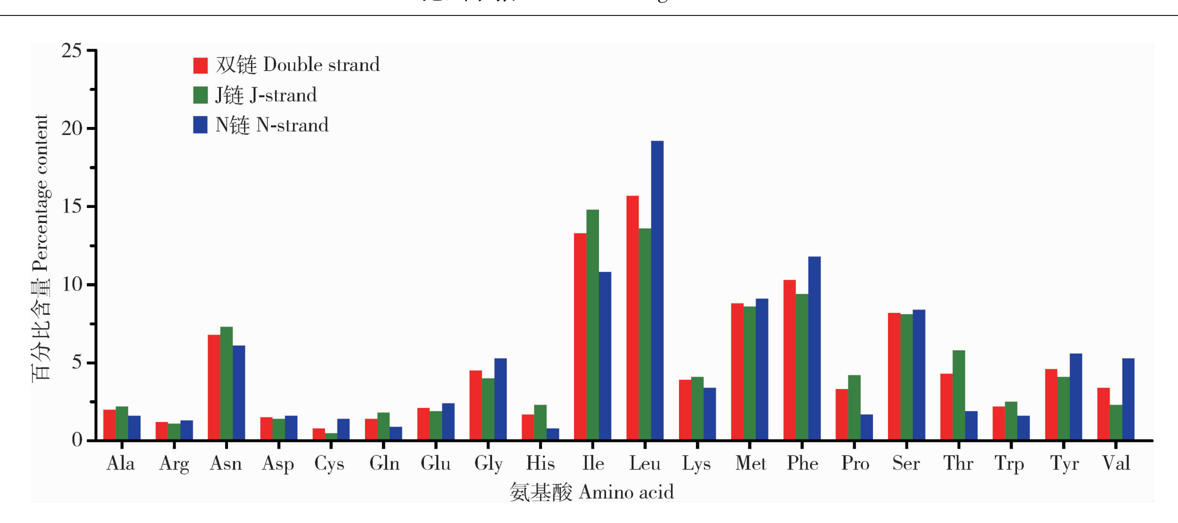


图 2　黄侧异腹胡蜂线粒体基因组蛋白质编码基因编码蛋白的氨基酸百分比含量

Fig. 2　Percentage of each amino acid of proteins coded by protein-coding genes of the mitochondrial genome of *Parapolybia crocea*

2016)，除了矮马蜂 *Polistes humilis* 中 $tRNA^{Ala}$ 的反密码子有所不同、为 AGC 外，目前已报道胡蜂科种类与黄侧异腹胡蜂所有 tRNA 基因的反密码子均相同(Cameron *et al*., 2008)。

除 $tRNA^{Ser(AGN)}$ 缺少的双氢尿嘧啶(DHU)臂被一个环结构所替代外，其他 tRNA 预测的二级结构均能折叠成典型的三叶草结构，包括氨基酸接受臂、双氢尿嘧啶臂、反密码子臂、可变臂以及 TψC 臂(图 3)。在三叶草结构的折叠过程中出现 19 处碱基错配，分别是 G-U 碱基错配 12 对，5 对存在于氨基酸接受臂上，6 对存在于双氢尿嘧啶臂上，1 对存在于 TψC 臂上；A-C 碱基错配 2 对，分别出现在氨基酸接受臂和反密码子臂上；发生在双氢尿嘧啶臂上的 G-A 碱基错配和氨基酸接受臂上的 U-U 碱基错配各 1 对。

与大部分昆虫线粒体基因组相比(Clary and Wolstenholme, 1985; Jiang *et al*., 2016)，黄侧异腹胡蜂线粒体基因组的 *rrnS* 基因位置相同，在 $tRNA^{Val}$ 基因和控制区之间，相对保守，而 *rrnL* 基因位置有所不同，位于蛋白质编码基因 *ND*1 和 $tRNA^{Val}$ 基因之间，*rrnL* 和 *rrnS* 长度分别为 1 424 和 752 bp，A + T 含量分别为 86.4% 和 85.9%，具有明显的 AT 碱基偏向性。

2.5　黄侧异腹胡蜂线粒体基因组控制区

黄侧异腹胡蜂线粒体基因组控制区是其基因组中序列最长的非编码区，位于 *rrnS* 和 $tRNA^{Tyr}$ 基因之间，长度为 1 316 bp，A + T 含量为 88.3%，表现出明显的 AT 碱基偏向性；控制区上存在一段保守的长为 18 bp 的 T-stretch 结构和 2 段串联重复序列(图 4)，第一段重复单位为 48 bp，重复 11 次有余，总长度 562 bp，位于基因组的 15 259 ~ 820 bp 之间；第 2 段重复单位为 41 bp，重复 2 次有余，总长度共 116 bp，位于基因组的 16 086 ~ 16 201 bp 之间，两段串联重复序列之间存在 292 bp 的基因间隔。

2.6　黄侧异腹胡蜂在胡蜂科中的系统发育地位

以蚁科的黑火蚁为外群，选用已报道的 6 个胡蜂科种类和本研究中测得的黄侧异腹胡蜂的线粒体基因组，基于 13 个蛋白质编码基因的核苷酸序列，用 PAUP 4.0 和 MEGA 6 预测出的最优模型均为 GTR + G + I，同时采用最大似然法(ML)构建胡蜂科的系统发育树，两种软件构建的系统发育树拓扑结构一致，均具有很高的置信值(图 5)。

结果显示，在胡蜂科中，马蜂亚科与胡蜂亚科聚为一支，3 个亚科之间的系统发育关系表现为蜾蠃亚科 + (胡蜂亚科 + 马蜂亚科)，亚科之内各属均明显分开，置信值高，分支的可信度较高。

3　讨论

与推测的昆虫祖先序列相比，膜翅目昆虫的线粒体基因组序列普遍存在基因重排现象，其中广腰亚目的基因排列相比细腰亚目保守，重排较少，但仍比其他目昆虫重排频繁(Jermiin and Crozier, 1994)。胡蜂科属于细腰亚目，重排事件主要发生在 tRNA 基因中，蛋白质编码基因和 rRNA 基因的排列顺序均相对保守，并且胡蜂科主要的重排热点发生在基因簇 *trnI*-*trnY*，*trnR*-*ND*5 和 *ND*1-*trnL*1 之间。在马蜂亚科矮马蜂 *Polistes humilis* 和约马蜂 *P. jokahamae* 已获得的线粒体基因序列中，主要存在两处基因重排现象，分别是 tRNA 基因 $tRNA^{Tyr}$ 和

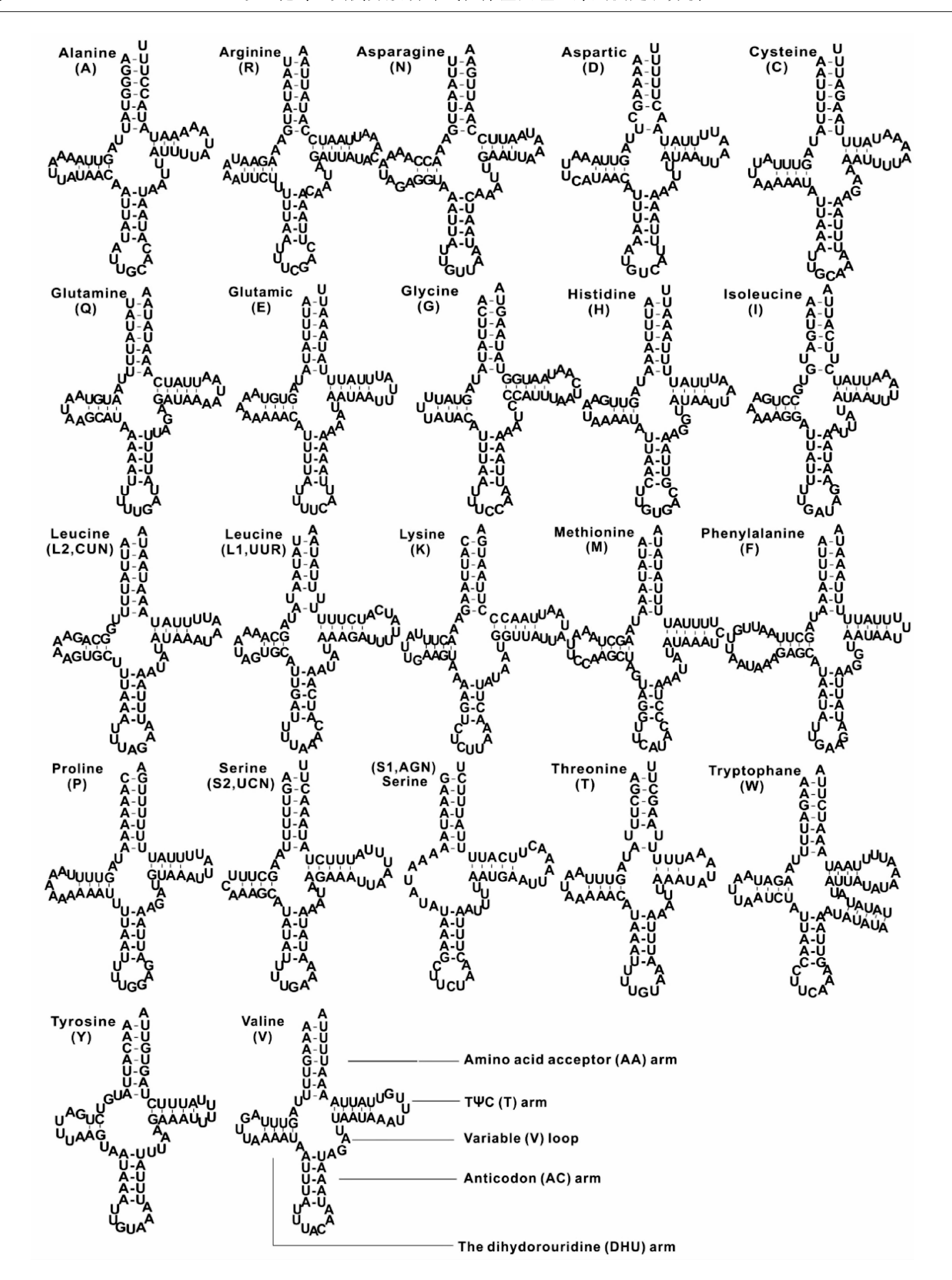


图 3　黄侧异腹胡蜂线粒体基因组 tRNA 基因产物的二级结构

Fig. 3　Predicted secondary structure of tRNA in the mitochondrial genome of *Parapolybia crocea*

$tRNA^{Leu(CUN)}$的重新排列(Cameron *et al*., 2008; Song *et al*., 2016a),本研究测定的黄侧异腹胡蜂线粒体基因组与已获得的马蜂亚科线粒体基因组的重排事件基本一致;胡蜂亚科获得完整线粒体基因组序列的种类金环胡蜂 *Vespa mandarinia* 中存在 4 处基因重排,包括$tRNA^{Tyr}$基因的移位,基因簇 *trnI*-*trnQ*-

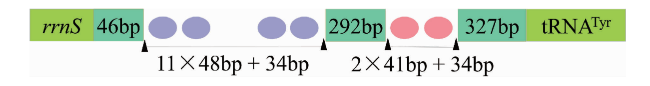


图 4 黄侧异腹胡蜂线粒体基因组控制区结构

Fig. 4 Structure of the control region in the mitochondrial genome of *Parapolybia crocea*

紫色圆表示第 1 段重复单位为 48 bp，粉色圆表示第 2 段重复单位为 41 bp。The purple circle denotes the first repeat unit (48 bp), while the pink circle denotes the second (41 bp).

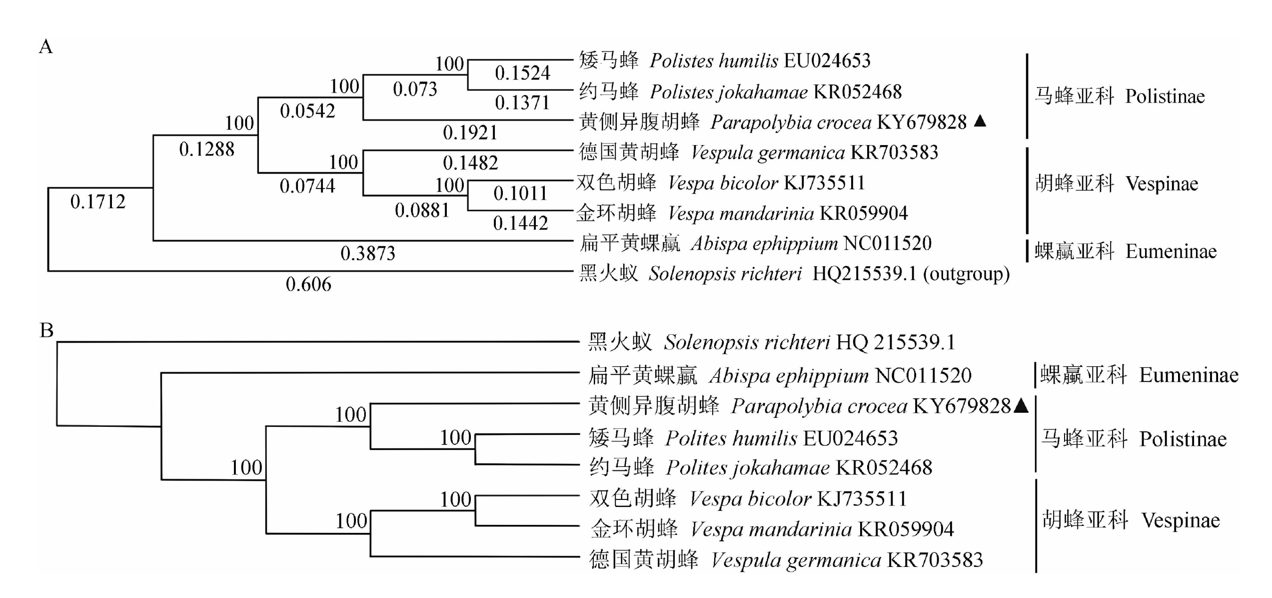


图 5 胡蜂科 7 个种基于线粒体 13 个蛋白质编码基因核苷酸序列的系统发育树（最大似然法）

Fig. 5 Phylogenetic tree of Vespidae based on the nucleotide sequences of 13 protein coding genes of mitochondrial genomes of seven species (maximum likelihood method)

A：使用 MEGA 6 构建的系统发育树 Phylogenetic tree constructed using MEGA 6；B：使用 PAUP 4.0 构建的系统发育树 Phylogenetic tree constructed using PAUP 4.0. 以黑火蚁 *Solenopsis richteri* 为外群。*Solenopsis richteri* was used as the outgroup.

trnM 的重新排列，*trnS*1 和 *trnE* 之间易位以及 *ND*1 和 *trnL*1 之间的易位（Chen *et al*., 2015），双色胡蜂 *V. bicolor* 和德国黄胡蜂 *Vespula germanica* 报道的线粒体基因组不完全，仅在获得的线粒体基因中发现存在部分与金环胡蜂类似的重排现象（Wei, 2014; Zhou *et al*., 2016）；蜾蠃亚科目前仅报道的扁平黄蜾蠃 *Abispae phippium* 中重排事件比较频繁，该线粒体基因组包括 41 个基因，其中有 2 个 $tRNA^{Met}$ 基因和 4 个 $tRNA^{Leu(UUR)}$ 基因，至少存在 4 次基因重排事件，形成了基因簇 *trnL*1-*trnM*-*trnQ*-*trnM*-*trnI*, *trnL*2-*trnL*2-*trnL*2-*trnL*2 和 *trnS*2-*ND*1（Cameron *et al*., 2008）。因此，从线粒体基因重排特点判断，异腹胡蜂属归为马蜂亚科有其合理性。

黄侧异腹胡蜂的 $tRNA^{Ser(AGN)}$ 的二级结构缺少双氢尿嘧啶臂而被一个长度为 7 nt 的环结构所替代，该现象是后生动物线粒体基因组 tRNA 的一个特征结构（Wolstenholme, 1992）。该种 tRNA 二级结构中存在保守区域，分别为 7 bp 长的氨基酸接受臂和 7 nt 长的反密码子环，其余长度可变的区域一般包括双氢尿嘧啶臂（2～4 bp）、双氢尿嘧啶环（3～8 nt）、反密码子臂（4～5 bp）、可变环（4～5 nt）、TψC 臂（3～5 bp）和 TψC 环（3～9 nt），其中 $tRNA^{Trp}$ 基因的可变环中间存在 1 个 5 bp 长的臂，$tRNA^{Phe}$ 基因的双氢尿嘧啶环长度达到了 13 nt，使得它们分别成为了长度最长的两个 tRNA 基因，这在目前报道的胡蜂科种类中尚未发现，需要在后续异腹胡蜂属相关种类的线粒体基因组研究中进一步收集并分析，判断其是否为该属线粒体基因组特有特征。

控制区是线粒体基因组中最长的一段非编码区，一般位于昆虫线粒体基因组的 *rrnS* 和 $tRNA^{Ile}$ 基因之间，是线粒体基因组中序列和长度变异最大的区域，除了核苷酸之间的替换，还包括序列的缺失插入，不同程度的串联重复拷贝等，通常认为控制区含有线粒体基因组转录和复制的起点（Saito *et al*., 2005）。黄侧异腹胡蜂的线粒体基因组控制区则位于基因组上的 *rrnS* 和 $tRNA^{Tyr}$ 基因之间，是目前报道

的胡蜂科线粒体基因组中控制区最长的种类。在其控制区上发现了一段长度为 18 bp 的 T-stretch 结构，有人推测该结构是完全变态昆虫线粒体基因组 N 链复制起始的结构信号（Saito *et al*., 2005），但是在胡蜂科其他种类控制区中未发现长度大于 10 bp 的 T-stretch 结构。黄侧异腹胡蜂控制区中还存在 2 段重复单位分别为 48 和 41 bp 的串联重复序列，在胡蜂科其他种类控制区中也发现不同程度的串联重复结构，说明胡蜂科线粒体基因的控制区具有更加多样化的结构特征。

基于胡蜂科 7 个种的线粒体基因组蛋白质编码基因的核苷酸序列构建的 ML 系统发育树分析可以看出，它们之间的系统发育关系为（（*Polistes humilis* + *P. jokahamae*）+ *Parapolybia crocea*）+（（*Vespa bicolor* + *V. mandarinia*）+ *Vespula germanica*）+ *Abispa ephippium*，马蜂亚科与胡蜂亚科的亲缘关系相较于蜾蠃亚科更近，这在形态与分子方面得到多方面的支持（Carpenter, 1982; Carpenter and Thi, 2003）；黄侧异腹胡蜂属与马蜂属的系统发育地位一致，同属于马蜂亚科，这与现沿用的形态分类学观点相一致（Saito-Morooka, 2015）。

本研究对黄侧异腹胡蜂线粒体基因组全序列的测定丰富了胡蜂科线粒体基因组的序列信息，为其分子系统进化分析提供了有益的分子数据；同时，对其结构特点和碱基组成的分析为对该属的线粒体基因组学进行比较分析提供了基础资料；最后，根据黄侧异腹胡蜂和胡蜂科 6 个线粒体基因组序列已知种的 13 个蛋白质编码基因核苷酸序列构建的系统发育树分析为该属分类地位的确立提供了线粒体基因组水平上的分子依据，也为后续胡蜂科系统发生关系的构建初步奠定了分子基础。但目前胡蜂科线粒体基因组全序列的数据仍有限，后续需要进一步对胡蜂科其他属的种类进行线粒体基因组测序和深入分析，从而构建较合理的胡蜂科各阶元间的进化关系。

致谢　感谢中国农业大学昆虫学系彩万志教授、李虎副教授及宋凡博士对本实验的大力支持。

参考文献（References）

Boore JL, 1999. Animal mitochondrial genomes. *Nucleic Acids Res*., 27(8): 1767–1780.

Brothers DJ, 1999. Phylogeny and evolution of wasps, ants and bees (Hymenoptera, Chrysidoidea, Vespoidea and Apoidea). *Zool. Scr*., 28(1–2): 233–250.

Cameron SL, 2014. Insect mitochondrial genomics: implications for evolution and phylogeny. *Ann. Rev. Entomol*., 59(1): 95–117.

Cameron SL, Dowton M, Castro LR, Ruberu K, Whiting MF, Austin D, Diement K, Stevens J, 2008. Mitochondrial genome organization and phylogeny of two vespid wasps. *Genome*, 51(10): 800–808.

Carpenter JM, 1982. The phylogenetic relationships and natural classification of the Vespoidea (Hymenoptera). *Syst. Entomol*., 7: 11–38.

Carpenter JM, Thi NLP, 2003. Keys to the genera of social wasps of South-East Asia (Hymenoptera: Vespidae). *Entomol. Sci*., 6(3): 183–192.

Carpenter, JM, 1993. Biogeographic patterns in the Vespidae (Hymenoptera): two views of Africa and South America. In: Goldblatt P ed. Biological Relationships between Africa and South America. Yale University Press, New Haven. 139–155.

Cha SY, Yoon HJ, Lee EM, Yoon MH, Hwang JS, Jin BR, Han YS, Kim I, 2007. The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae). *Gene*, 392(1–2): 206–220.

Chen PY, Wei SJ, Liu JX, 2015. The mitochondrial genome of the *Vespa mandarinia* Smith (Hymenoptera: Vespidae: Vespinae) and a phylogenetic analysis of the Vespoidea. *Mitochondr. DNA*, 27(6): 4414–4415.

Clary DO, Wolstenholme DR, 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol*., 22(3): 252–271.

Dowton M, Austin AD, 1997. The evolution of strand-specific compositional bias. A case study in the hymenopteran mitochondrial 16S rRNA gene. *Mol. Biol. Evol*., 14(1): 109–112.

Gotzek D, Clarke J, Shoemaker D, 2010. Mitochondrial genome evolution in fire ants (Hymenoptera: Formicidae). *BMC Evol. Biol*., 10(1): 300.

Hua YQ, Din YR, Yan ZT, Si FL, Luo QC, Chen B, 2016. The complete mitochondrial genome of *Anopheles minimus* (Diptera: Culicidae) and the phylogenetics of known *Anopheles* mitogenomes. *Insect Sci*., 23(3): 353–365.

Jermiin LS, Crozier RH, 1994. The cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger*: sequence divergence in Hymenoptera may be associated with nucleotide content. *J. Mol. Evol*., 38(3): 282–294.

Jiang P, Li H, Song F, Cai Y, Wang JY, Liu JP, Cai WZ, 2016. Duplication and remolding of tRNA genes in the mitochondrial genome of *Reduvius tenebrosus* (Hemiptera: Reduviidae). *Int. J. Mol. Sci*., 17(6): 951.

Lee YS, Oh J, Kim YU, Kim N, Yang S, Hwang UW, 2008. Mitome: dynamic and interactive database for comparative mitochondrial genomics in metazoan animals. *Nucleic Acids Res*., 36: 938–942.

Li TS, 1982. Hornets from Agricultural Regions of China. China Agriculture Press, Beijing. 98–103. [李铁生，1982. 中国农区胡蜂. 北京：农业出版社. 98–103]

Li TS, 1985. Economic Insect Fauna of China, Fasc. 30. Hymenoptera:

Vespoidea. Science Press, Beijing. 54 – 58. [李铁生, 1985. 中国经济昆虫志, 第30册. 膜翅目: 胡蜂总科. 北京: 科学出版社. 54 – 58]

Luo QC, Hao YJ, Meng FX, Li TJ, Ding YR, Hua YQ, Chen B, 2016. The mitochondrial genomes of *Culex tritaeniorhynchus* and *Culex pipiens pallens* (Diptera: Culicidae) and comparison analysis with two other *Culex* species. *Parasit. Vectors*, 9: 406.

Peng Y, Leung HCM, Yiu SM, Chin FYL, 2012. IBDA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28: 1420 – 1428.

Richards OW, 1962. A Revisional Study of the Masarid Wasps (Hymenoptera, Vespoidea). British Museum (Natural History), London. 294 pp.

Saito S, Tamura K, Aotsuka T, 2005. Replication origin of mitochondrial DNA in insects. *Genetics*, 171(4): 1695 – 1705.

Saito-Morooka F, Nguyen LT, Kojima J, 2015. Review of the paper wasps of the *Parapolybia indica* species-group (Hymenoptera: Vespidae, Polistinae) in eastern parts of Asia. *Zootaxa*, 3497(2): 215 – 235.

Saussure H, 1853 – 1858. études sur la Famille des Vespides. 2. Monographie des Guêpes Sociales, ou de la Tribu des Vespiens. V. Masson, Paris and J. Kessmann, Genève. 398 – 460.

Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT, 2006. Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.*, 37(1 – 2): 545 – 579.

Song SN, Chen PY, Wei SJ, Chen XX, 2016a. The mitochondrial genome of *Polistes jokahamae* and a phylogenetic analysis of the Vespoidea (Insecta: Hymenoptera). *Mitochondr. DNA*, 27(4): 1 – 2.

Song SN, Tang P, Wei SJ, Chen XX, 2016b. Comparative and phylogenetic analysis of the mitochondrial genomes in basal hymenopterans. *Sci. Rep.*, 6: 20972.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.*, 30(12): 2725 – 2729.

Vecht vd J, 1966. The East-Asiatic and Indo-Australian species of *Polybioides* Buysson and *Parapolybia* Saussure (Hym., Vespidae). *Zool. Verhandel.*, 82: 1 – 42.

Wang JP, Xuan SB, Zhang YP, Yang J, Cao TW, Ma EB, 2015. Sequencing and analysis of the complete mitochondrial genome of *Chitoria ulupi* (Lepidoptera: Nymphalidae). *Acta Entomol. Sin.*, 58(3): 319 – 328. [王菊平, 宣善滨, 张育平, 杨静, 曹天文, 马恩波, 2015. 武铠蛱蝶线粒体基因组全序列测定和分析. 昆虫学报, 58(3): 319 – 328]

Wei SJ, Min S, He JH, Sharkey M, Chen XX, 2009. The complete mitochondrial genome of *Diadegma semiclausum* (Hymenoptera: Ichneumonidae) indicates extensive independent evolutionary events. *Genome*, 52(4): 308 – 319.

Wei SJ, Niu FF, Tan JL, 2014. The mitochondrial genome of the *Vespa bicolor* Fabricius (Hymenoptera: Vespidae: Vespinae). *Mitochondr. DNA*, 27(2): 1 – 2.

Wei SJ, Shi M, Chen XX, Sharkey MJ, van Achterberg C, Ye GY, He JH, 2010. New views on strand asymmetry in insect mitochondrial genomes. *PLoS ONE*, 5(9): e12708.

Wolstenholme DR, 1992. Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.*, 141(6): 173 – 216.

Zhao GY, Li H, Zhao P, Cai WZ, 2015. Comparative mitogenomics of the assassin bug genus *Peirates* (Hemiptera: Reduviidae: Peiratinae) reveal conserved mitochondrial genome organization of *P. atromaculatus*, *P. fulvescens* and *P. turpis*. *PLoS ONE*, 10: e0117862.

Zhou Y, Hu YL, Xu ZF, Wei SJ, 2016. The mitochondrial genome of the German wasp *Vespula germanica* (Fabricius, 1793) (Hymenoptera: Vespoidea: Vespidae). *Mitochondr. DNA*, 27(4): 1 – 2.

（责任编辑：马丽萍）